This is the peer reviewed version of the following article:
Mitsudo, H. (2004). Rapid image-segmentation and perceptual transparency share a process which utilises X-junctions generated by temporal integration in the visual system. *Perception, 33(4),* 471-484,
which has been published in final form at http://dx.doi.org/10.1068/p5214.

Running head: TRANSPARENCY AND TEMPORAL INTEGRATION

Rapid image-segmentation and perceptual transparency share a process which utilizes X-junctions generated by temporal integration in the visual system

Hiroyuki Mitsudo
Kyushu University

Word count:
Main text: 4,935
Total: 6,817

Correspondence:
Hiroyuki Mitsudo
Department of Psychology, Faculty of Letters, Kyushu University,
6-19-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
Phone/Fax: +81926422418
E-mail: hmitsudo@lit.kyushu-u.ac.jp

Abstract

Perceptual transparency requires local same-polarity X-junctions, which can also be generated by temporal integration under natural dynamic conditions. In this study, segmentation performance and target appearance were measured for a uniform gray target embedded in a random-dot frame presented with a temporally adjacent mask. Although static cues for both segmentation and transparency were unavailable, transparency was observed only when collinear same-polarity edges reduced backward masking, in both the fovea and the perifovea. These results suggest that the visual system has a common underlying mechanism for rapid segmentation and transparency, which utilizes same-polarity X-junctions generated by temporal integration.

1 Introduction

Visual information-processing is frequently assumed to consist of two functionally distinctive stages (Julesz 1991; Neri and Heeger 2002; Scialfa and Joffe 1995), sometimes termed as segmentation and identification. The former stage involves segmenting an input image into several regions, by spatially comparing only local ‘differences’ in each primitive feature such as luminance (Forte et al 1999) and orientation (Motoyoshi and Nishida 2001). The latter stage involves identifying detailed properties (eg shape and reflectance) of a spatially localized segment. Studies using a backward-masking paradigm suggest that such ‘two-dimensional’ segmentation is a rapid process: even when a target defined by such features is briefly presented (for less than 100 ms) at an unpredictable location and then followed by another mask pattern, segmentation performance is quite good (Sagi and Julesz 1985; Gegenfurtner and Rieger 2000). In contrast, identification is a slower, time-consuming process (Neri and Heeger 2002; Scialfa and Joffe 1995).

This framework, assuming a strict distinction between segmentation and identification, is questioned by psychophysical evidence regarding perceptual transparency. That is, although transparency is traditionally considered as reflecting how the visual system not only ‘segments’ stratified surfaces[1] but also ‘identifies’ perceived shape, reflectance and transmittance of segmented regions that are embedded in a two-dimensional mosaic image (Metelli 1974; Kersten 1991; Singh and Anderson 2002), several studies indicate that segmentation accompanying transparency is rapid (Watanabe and Cavanagh 1992; Mitsudo 2003). For example, Watanabe and Cavanagh (1992) showed that transparency requires an exposure duration of only about 60 ms, using a pattern recognition task; Mitsudo (2003) showed that the ‘detailed’ properties (such as shape and reflectance) accompanied by transparency influence rapid search[2]. According to these studies, rapid segmentation in transparency is primarily constrained by relatively complex luminance-defined static features, known as ‘same-polarity’ X-junctions (Figure 1A). The junctions are defined as intersections at which contours of an object not only cross the background contour, but also preserve the contrast polarity of the background contour (Figures 1A and 1B, Adelson and Anandan 1990; Anderson 1997; Beck et al 1984).

From an ecological viewpoint, however, it is unclear why the visual system has the rapid segmentation process which utilizes same-polarity X-junctions. That is, in a natural environment such static X-junctions, which are typically generated by clear sheets, mesh and haze, seem to be much less frequent[3] than other static local cues, such as T-junctions (Figure 1C) for occlusion.

This problem can be solved by the idea that transparency shares an underlying mechanism with the rapid image-segmentation process that compensates for temporal integration (ie signal averaging) in the visual system. In formulating a quantitative model of transparency, temporal integration was conceptually introduced by Metelli (1974). The central idea of his episcotister model is that the luminance profile of regions composing a transparent surface can be predicted by the mixture of the reflectance of a fast-moving opaque object and that of its static patterned background. By assuming that temporal integration makes the fast-moving opaque object perceptually equivalent to a transparent surface, this model successfully describes the above-mentioned constraint on transparency (ie same-polarity X-junctions). On the other hand, under natural 'dynamic' conditions, temporal integration in the visual system is not negligible especially at a short temporal range, since visual response to a briefly presented stimulus persists for about 50 ms after its offset (Dixon and Di Lollo 1994). That is, whenever an observer fixates on a patterned background, continuous fast movement of an object always produces 'streaks' that frequently intersect with contours of the background (Figure 1D). Accordingly, same-polarity X-junctions can be frequent visual features accompanied by dynamic objects, and therefore can be a constraint on segmentation under dynamic conditions. This constraint is computationally parallel to the case of transparency under static conditions. It is thus parsimonious to presume a common mechanism for transparency and segmentation under dynamic conditions, which utilize same-polarity X-junctions.

===========================
Insert Figure 1 about here
===========================

The purpose of the present study was to examine the hypothesis that perceptual transparency is closely related to the rapid segmentation process that utilizes same-polarity X-junctions generated by temporal integration. Specifically, I investigated how spatio-temporal factors influence two-dimensional segmentation and a transparency judgment[4], and sought a tight connection between them. Holcombe (2001) recently reported that temporal integration can trigger transparency in the absence of static X-junctions, by using temporal alteration of spatial gratings with different orientations. His claim contradicts the above hypothesis: he argued that this 'temporal' transparency is related to a high-level attentional process that integrates the visual signal across about 120 ms. However, he did not examine segmentation performance; in addition, his periodic presentation may be inadequate for investigating the elementary segmentation process, since the visual system is highly sensitive to the onset of periodic stimuli (Beaudot 2002).

In the present experiment, observers were required to report both the location (segmentation task) and the appearance (identification task) of a target which appeared at an unpredictable location in one of two random-dot frames that were presented briefly in succession (Figure 2A). In the identification task, three categories ('transparent', 'mosaic' and 'opaque') were used (Figure 2C). To investigate the effects of temporal integration on the formation of X-junctions, I used 'asynchronous' X-junctions, from which static cues for both segmentation and transparency had been eliminated. These junctions were defined by the dot relation of the two frames around the contour of a physically uniform target gray bar (Figure 3). By varying contrast polarity at these junctions, I was able to introduce same- and opposite-polarity X-junctions, whose difference in performance provided a measure of the advantage of asynchronous same-polarity X-junctions in segmentation and identification.

The rapid segmentation process can be measured with performance improvement in backward masking (Gegenfurtner and Rieger 2000). Backward masking, but not forward masking, is powerful for disrupting the processing of a preceding target, especially when the target is presented at an unpredictable location (Breitmeyer and Ogmen 2000; Enns and Di Lollo 1997). To ascertain that segmentation performance is selectively disrupted by backward masking, I varied both the temporal order and the exposure duration of the two frames (total duration ranged from 40 to 120 ms). Since half elements of X-junctions were presented at an unpredictable location within the preceding frame, the masking theories predict no segmentation advantage of such asynchronous same-polarity X-junctions for any temporal conditions.

The above hypothesis -- transparency is closely related to the rapid segmentation process that utilizes same-polarity X-junctions generated by temporal integration -- implies both that (a) rapid segmentation can utilize asynchronous same-polarity X-junctions, and that (b) transparency[5] is closely related to this segmentation process. To test (a), I examined whether or not asynchronous same-polarity X-junctions can improve segmentation accuracy when backward masking (rather than forward masking) occurs. This is because the rapid segmentation process is thought to reduce backward masking especially at a short stimulus onset asynchrony (SOA). To test (b), I examined whether or not the improvement in segmentation accuracy is attributed only to the increase of the ‘transparent’ responses in the identification task. This is because in the above hypothesis, both rapid segmentation and transparency utilize the same information -- same-polarity X-junctions generated by temporal integration.

I tested these predictions with a target presented in the fovea and the perifovea of the visual field, in separate groups (foveal and perifoveal groups, respectively). Rapid segmentation is thought to be essentially a spatially parallel process (Sagi and Julesz 1985). Thus if the above hypothesis is correct, then the predictions will be supported for both groups.

==========================

Insert Figure 2 about here

==========================

2 Methods

2.1 Observers

Fifteen students (female, seven; male, eight) of Kyushu University participated in the experiment. Nine of the observers were assigned to the foveal group, while the other six were assigned to the perifoveal group. All had normal or corrected-to-normal vision according to self-reporting, and all were naive to the purpose of the experiment. Before commencing the experimental session, all the observers were informed about the procedure of the experiment and the duration of the experimental session (about 40 minutes).

2.2 Apparatus

The experiment was conducted on an Apple iMac DV with a CRT monitor (1,024 x 768 pixels; vertical refresh rate, 75 Hz) with gamma correction. Stimulus generation and data collection were controlled by a program written in C.

2.3 Stimuli

The stimuli comprised both target and mask frames (visual angle, 2.8° x 2.8° for the foveal group; 16.9° x 16.9° for the perifoveal group), each of which consisted of 8 x 8 binary random dots (each dot subtended 0.35° x 0.35° for the foveal group and 2.11° x 2.11° for the perifoveal group), and which were presented in succession (Figure 2A). The target frame contained either a vertical or a horizontal uniform gray bar (visual angle, 2.1° x 0.7° for the foveal group; 12.7° x 4.2° for the perifoveal group) that was presented at one of four possible locations (Figure 2B). The mask frame did not contain the target. For each trial, the bright and dark dots constituting the two frames were randomly placed with equal probabilities (bright:dark = 1:1), except for those around the corners of the target in an effort to avoid ambiguous X-junctions, and those of the mask frame within the target location (see next paragraph). The luminance of the bright and dark dots was 40 and 20 cd/m^2, respectively. The luminance of the gray bar target and the blank display was 30 cd/m^2 (the average luminance of the bright and dark dots). A small black fixation dot (visual angle, 0.05° x 0.05° for the foveal group; 0.10° x 0.10° for the perifoveal group) was constantly present at the center of the screen during an observation sequence. Stimuli for each trial were prepared on the computer memory during the presentation of the fixation at the beginning of the trial.

There were two contrast polarity conditions, same- and opposite-polarity conditions, which differed only in terms of the dot relation between the target and mask frames around the target contour (Figure 3). In the same-polarity condition, dots inside the target location in the mask frame were arranged so as to preserve the contrast polarity of the dots around the target contour in the target frame. In the opposite-polarity condition, conversely, the dots of the mask frame were reversed against those of the target frame. Because the other dots of the two frames were randomly distributed and because the target was always a uniform gray bar, no static cue for transparency[6] or segmentation was available. In addition, no motion was perceived between these relevant dots.

============================

Insert Figure 3 about here

============================

2.4 Procedure

Observers viewed the monitor in a binocular fashion, at a distance of 60 cm in a darkened room, with the head stabilized on a chin rest. Through each observation sequence, observers were required to fixate on the small black dot at the center of the screen. Each observation sequence began with a blank screen presented for 500 ms; after the sequential presentation of the two random-dot frames, the blank screen was again presented for 500 ms (Figure 2A). The target and mask frames had the same duration for each trial, with the exposure duration of 13, 27 or 53 ms (eg 13 ms for the target frame and 13 ms for the mask frame, Figure 2D). To avoid ‘optical’ integration between the two frames caused by phosphor persistence of a CRT monitor, a blank frame was presented for 13 ms between the two frames[7].

After each observation sequence, observers performed both segmentation and identification tasks. First, observers were required to report the location at which a ‘grayish’ target appeared (segmentation task) based on a spatial four-alternative forced choice. The response display for this task, containing white outlines of four possible target locations (Figure 2B), was presented in the same size as the two random-dot frames at the center of the screen.

Second, observers were required to report the appearance of the target (identification task) based on a three-alternative forced choice. These three response categories were ‘transparent’, ‘mosaic’ and ‘opaque’, which were determined according to preliminary observations. Observers were asked to choose one of the three categories according to their perceptual criteria without considering the balance among the three categories when pressing the keys. None of the observers were told that the target was always a physically uniform gray bar. The response display for the identification task consisted of three patterns corresponding with the three response categories (Figure 2C). Each pattern was a modification of the target frame for each trial (therefore, these patterns also served as a feedback for the segmentation task). These patterns were aligned horizontally; they were presented below the observation display for the foveal group and presented at the center of the monitor for the perifoveal group. For both groups, each pattern had the same size as that of the target frame for the foveal group. In the patterns for the ‘transparent’ and ‘mosaic’ responses the luminance of the target region was varied according to Metelli’s rule. That is, the pattern for the ‘transparent’ response (Figure 2C, left panel) had static same-polarity X-junctions which satisfy the luminance condition for transparency (Metelli 1974; Singh and Anderson 2002), whereas the pattern for the ‘mosaic’ response had static opposite-polarity X-junctions which did not satisfy the luminance condition (Figure 2C, center panel). The luminance of the darker and lighter dots presented within the target region was 25 and 35 cd/m^2, respectively. The pattern for the ‘opaque’ response was identical to that of the target frame (Figure 2C, right panel).

For each task, observers responded by pressing the ‘up’, ‘down’, ‘left’ or ‘right’ key on the keyboard with their right hand. The current state of a chosen response was indicated with a thick white outline (line thickness, 0.06° for the foveal group; 0.12° for the perifoveal group). The response display for each task was visible until observers pressed the ‘space’ key; the second key press (for the identification task) triggered the next trial. Observers were not required to make speedy judgments.

Each observer underwent 288 trials separated into six blocks. In each block (consisting of 48 trials), all combinations of all conditions, (target location, 4; contrast polarity, 2; presentation order of the two frames, 2; duration of each frame, 3) were presented once in randomized order. The first block served as practice trials, and was excluded from the data analysis. The locations of the stimuli corresponding to the three categories for the identification task were counterbalanced across observers.

3. Results and Discussion

Results were analyzed as a function of SOA, which combines the two temporal factors (presentation order and exposure duration, Figure 2D). A positive SOA means that the target appeared in the first frame; a negative SOA means that the target appeared in the second frame. Absolute values of SOA (27, 40 and 67 ms) correspond to the exposure duration (13, 27 and 53 ms, respectively).

3.1 Foveal group

3.1.1 Segmentation task. Figure 4A shows the mean accuracy for the segmentation task as a function of SOA, averaged over the nine observers. Two-way within-observers analysis of variance (ANOVA) was performed on the arcsine-transformed accuracy data, with the factors of SOA (–67, –40, –27, +27, +40, +67) and contrast polarity at the X-junctions (same, opposite). The main effects of SOA and contrast polarity were significant [$F(5, 40) = 138.10$, $p < .0001$; $F(1, 8) = 30.57$, $p < .001$, respectively], and the interaction was also significant, $F(5, 40) = 7.46$, $p < .0001$. Furthermore, at SOAs of –27, +27 and +40 ms, accuracy was significantly higher in the same-polarity condition than in the opposite condition [$F(1, 40) = 14.99$, $p < .0005$; $F(1, 40) = 30.48$, $p < .0001$; $F(1, 40) = 27.90$, $p < .0001$, respectively].

==============================

Insert Figure 4 about here

================================

3.1.2 Identification task. The correct responses in the segmentation task were categorized into three types (transparent, mosaic and opaque) according to the judgment in the identification task.

Figures 4B-D show the mean correct frequency for each category as a function of SOA. The sum of the three responses is equal to the mean segmentation accuracy for each polarity condition. Dashed and dotted lines indicate the chance levels. Two-way within-observers ANOVA was performed on the arcsine-transformed correct responses for each category, with the factors of SOA (–67, –40, –27, +27, +40, +67) and contrast polarity of the X-junctions (same, opposite). The two-way interaction was significant for all categories [transparent: $F(5, 40) = 6.92$, $p < .0001$; mosaic: $F(5, 40) = 3.66$, $p < .01$; opaque: $F(5, 40) = 5.71$, $p < .001$]. A significant advantage of the same-polarity condition over the opposite-polarity condition was obtained for the ‘transparent’ responses at SOAs of +27, +40 and +67 ms [$F(1, 40) = 17.00$, $p < .0005$; $F(1, 40) = 5.50$, $p < .05$; $F(1, 40) = 4.93$, $p < .05$, respectively] and for the ‘opaque’ responses at SOAs of –67 and –40 ms [$F(1, 40) = 19.05$, $p < .0001$; $F(1, 40) = 31.30$, $p < .0001$, respectively]. An opposite-polarity advantage was obtained for the ‘transparent’ responses at an SOA of –67 ms [$F(1, 40) = 7.41$, $p < .01$] and for the ‘mosaic’ responses at SOAs of –67, –40 and +67 ms [$F(1, 40) = 5.65$, $p < .05$; $F(1, 40) = 4.56$, $p < .05$; $F(1, 40) = 7.50$, $p < .01$, respectively]. Asterisks in Figures 4B-D indicate these significant differences between the same- and opposite-polarity conditions.

3.2 Perifoveal group

3.2.1 Segmentation task. Figure 5A shows the mean accuracy for the segmentation task as a function of SOA, averaged over the six observers. Two-way within-observers ANOVA was performed on the arcsine-transformed accuracy data, with the factors of SOA (–67, –40, –27, +27, +40, +67) and contrast polarity of the X-junctions (same, opposite). The main effects of SOA and contrast polarity were significant [$F(5, 25) = 45.89$, $p < .0001$; $F(1, 5) = 11.18$, $\mathrm{p} < .05$, respectively], and the interaction was also significant, $F(5, 25) = 3.00$, $p < .05$. Furthermore, at SOAs of +27 and +40 ms, segmentation accuracy was significantly higher in the same-polarity condition than in the opposite-polarity condition [$F(1, 25) = 6.31$, $p < .05$; $F(1, 25) = 14.66$, $p < .001$, respectively].

==============================

Insert Figure 5 about here

==============================

3.2.2 Identification task. The correct responses in the segmentation task were categorized into three types (transparent, mosaic and opaque), according to the judgment in the identification task.

Figures 5B-D show the mean correct frequency for each category as a function of SOA. The sum of the three responses is equal to the mean segmentation accuracy for each polarity condition. Again, two-way within-observers ANOVA was performed on the arcsine-transformed correct responses for each category, with the factor of SOA (–67, –40, –27, +27, +40, +67) and contrast polarity of the X-junctions (same, opposite). The two-way interaction was significant for all categories [transparent: $F(5, 25) = 4.86$, $p < .005$; mosaic: $F(5, 25) = 3.83$, $p < .05$; opaque: $F(5, 25) = 6.86$, $p < .0005$]. A significant advantage of the same-polarity condition was obtained for the ‘transparent’ responses at SOAs of +27, +40 and +67 ms [$F(1, 25) = 16.29$, $p < .001$; $F(1, 25) = 6.68$, $p < .05$; $F(1, 25) = 4.37$, $p < .05$, respectively], and for the ‘opaque’ responses at SOAs of –67, –40, –27, +40 and +67 ms [$F(1, 25) = 27.50$, $p < .0001$; $F(1, 25) = 55.64$, $p < .0001$; $F(1, 25) = 5.06$, $p < .05$; $F(1, 25) = 4.79$, $p < .05$; $F(1, 25) = 13.70$, $p < .005$, respectively]. An opposite-polarity advantage was obtained for the ‘transparent’ responses at an SOA of –40 ms [$F(1, 25) = 5.98$, $p < .05$] and for the ‘opaque’ responses at SOAs of –67, –40 and +67 ms [$F(1, 25) = 19.72$, $p < .0005$; $F(1, 25) = 10.94$, $p < .005$; $F(1, 25) = 13.51$, $p < .005$, respectively]. Asterisks in Figures 5B-D indicate these significant differences between the two polarity conditions.

3.3 Discussion

The results of the segmentation task showed that asynchronous same-polarity X-junctions can improve segmentation accuracy at SOAs of –27 to +40 ms for the foveal group (Figure 4A) and at SOAs of +27 and +40 ms for the perifoveal group (Figure 5A), indicating that the X-junctions can reduce backward masking. This same-polarity advantage was attributed only or mainly to the increase in the ‘transparent’ (Figures 4B and 5B) but not either the ‘mosaic’ (Figures 4C and 5C) or ‘opaque’ (Figures 4D and 5D) responses in the identification task. Outside these SOAs, asynchronous X-junctions operated in a very different manner: the ‘transparent’ responses did not exceed the chance level (Figures 4B and 5B) although the inter-stimulus interval was constant at all SOAs; rather, temporal ‘separation’, measured with the ‘opaque’ responses, increased in the same-polarity condition, whereas it was rarely observed when backward masking occurred (Figures 4D and 5D). Thus, these results for both the foveal and perifoveal groups indicate that transparency is closely related to the rapid segmentation process utilizing same-polarity X-junctions generated by temporal integration.

Transparency observed only when backward masking occurred is different from response bias accompanied by a task difficulty at the SOAs. For example, one might think that the increase in ‘transparent’ responses could simply be due to a response bias for the ‘transparent’ category with which observers associated low visibility of the target, since the luminance contrast of the ‘transparent’ target in the identification display was lower than that of the other two stimuli. The visibility can be measured with the segmentation accuracy; thus, this explanation predicts that the ‘transparent’ responses would be more frequent in the opposite-polarity condition than in the same-polarity condition, since segmentation accuracy was lower in the opposite-polarity than in the same-polarity condition. However, the results obtained here exclude this explanation: as shown in Figures 4B and 5B, at these SOAs, the ‘transparent’ responses were less frequent in the opposite-polarity condition than in the same-polarity condition[8].

Furthermore, the same-polarity advantage in segmentation accuracy cannot be explained with low-spatial-frequency components accompanied by 'static' same-polarity X-junctions. One might think that, since in static transparency stimuli (such as the left panel in Figure 2C) the spatial frequency around the target region tends to be lower than that of the rest of the region, such low-spatial-frequency signals could contribute towards increasing the target visibility. If this explanation is correct, a same-polarity advantage in segmentation accuracy would be observed even when the target had no transparency. The results, however, ruled out this explanation: for the 'mosaic' responses (Figures 4C and 5C), the correct frequency in the same-polarity condition was nearly identical to that in the opposite-polarity condition.

One could argue that the same-polarity advantage accompanying transparency can be explained by previously reported local interactions between collinear edges. For example, Polat and Sagi (1993) reported that detection of a luminance-modulated Gabor target is facilitated by the surrounding collinear same-polarity edges, even if they have a spatio-temporal gap (Tanaka and Sagi 1998).

However, such local interactions are not sufficient for explaining the present data. First, such local interactions are highly dependent upon spatial configuration between the target and the surroundings, and are greatly reduced by the non-collinear surrounding edges (Solomon and Morgan 2000). Since I used a random-dot display which contained non-collinear surroundings, such interactions alone do not seem to explain the present results. Second, such lateral interactions are dependent upon target eccentricity (Williams and Hess 1998; Xing and Heeger 2000): the luminance modulation of a target is seen more easily with surrounding gratings than with a uniform background, when the target is presented not in the periphery but in the fovea. In the present experiment, target eccentricity was larger for the perifoveal group than for the foveal group. Thus, if the present results were due to the accidental by-product of such simple lateral interactions, then the same-polarity advantage would be observed only for the foveal group, and not for the perifoveal group. This prediction was not supported as shown in Figures 4 and 5: the results for the perifoveal group were very similar to those for the foveal group. Rather, since rapid segmentation is thought to be essentially a spatially parallel process (Sagi and Julesz 1985), the same-polarity advantage observed for both groups seems to reflect the rapid segmentation process.

4 General Discussion

The experiment clearly showed that transparency is closely related to the rapid segmentation process which utilizes same-polarity X-junctions generated by temporal integration. A parsimonious explanation is that this tight connection is caused by a common mechanism sensitive to same-polarity X-junctions, which can be a constraint on both segmentation under dynamic conditions and perceptual transparency. This is strengthened by the results that asynchronous same-polarity X-junctions did not facilitate segmentation when coded as 'mosaic' (Figures 4C and 5C). Thus, a strict distinction between segmentation and identification, simply based on local computation of primitive features, is not valid; the present results are consistent with a recent claim that the visual system is highly sensitive to features generated under natural settings, such as the spatial conjunction between luminance and color (Gegenfurtner and Rieger 2000) and motion streaks (Geisler 1999).

Furthermore, these results exclude a low-temporal-resolution account (Holcombe 2001) for temporal transparency. Holcombe (2001) claimed that the integration time for triggering temporal transparency is about 120 ms by using simultaneity rating. In the present study, I used a two-frame display with a constant inter-stimulus interval (13 ms) in which total duration was within 120 ms at all SOAs. If the account is correct, transparency would be observed at all SOAs. However, as shown in Figures 4B and 5B, the present results were not consistent with this account: transparency was observed only when same-polarity X-junctions reduced backward masking, although a simultaneous, but not transparent percept measured with the 'mosaic' responses was obtained with a relatively long duration (especially for the opposite-polarity condition, Figures 4C and 5C). These results indicate that the simultaneity rating may be inappropriate for measuring perceptual transparency.

This study is the first to demonstrate that same-polarity X-junction elements presented within about 50 ms can facilitate segmentation of a target which appears at an unpredictable location, even when the elements are asynchronously presented in a noisy sequence. This facilitatory effect can compensate for the essential difficulty in the visual system revealed by recent studies on visual masking. That is, detection of a target presented at an unpredictable location is severely impaired by the subsequent mask with a small temporal onset or offset (within about 100 ms), even when the target and the mask have a spatial gap (Enns and Di Lollo 1997; Jiang and Chun 2001). These studies cannot explain why such difficulty rarely seems to be experienced in natural dynamic settings. Although attentional mechanisms have been proposed to compensate for this difficulty (Shelley-Tremblay and Mack 1999), such mechanisms seem to be inadequate for covering the large visual field. The results of the segmentation task (Figures 4A and 5A) can provide an answer to this problem: in a natural dynamic setting, the high sensitivity of the visual system to same-polarity X-junctions can compensate for such a difficulty.

The same-polarity advantage obtained in this study could not be predicted by previous studies reporting the influence of contrast polarity (or phase) on detection of a luminance-modulated target (Georgeson 1988; Bowen and Wilson 1994; Foley and Chen 1999). For instance, Bowen and Wilson (1994) showed that detection of a Gabor target is facilitated by an overlapping sine-wave mask when the mask has the same phase as that of the target and the SOA is short (< 50 ms). This facilitation seems to correspond essentially to signal enhancement in luminance at the target location (Watson and Nachmias 1977), since in previous studies the target and the mask had a 'spatial' overlap. In the present study, I used a contrast-modulated target in which dots defining the polarity of the X-junctions did not spatially overlap with those of the mask; thus signal enhancement in luminance is insufficient for explaining the present results.

Given that transparency involves rapid segmentation, which of the properties accompanied by transparency (eg depth, shape, reflectance and transmittance) determines segmentation performance? Since transparency inevitably accompanies surface stratification (ie depth assignment) and apparent contrast suppression in the ‘stratified’ region, these properties may predict the same-polarity advantage in segmentation accuracy. This study does not provide a clear answer to this issue; that is, to establish a tight connection between rapid segmentation and temporal transparency, I focused on a transparency judgment, rather than a surface-stratification or contrast-suppression judgment. This is because either surface stratification or contrast suppression can be accompanied by other phenomena, such as amodal completion (Rauschenberger and Yantis 2001) or lateral interactions (Xing and Heeger 2000), respectively.

In contrast to other cues for perceptual transparency (eg binocular disparity and motion), same-polarity X-junctions generated by temporal integration are a robust cue for rapid segmentation. In the case of both binocular disparity (Akerstrom and Todd 1988) and motion (Masson et al 1999), transparency requires a relatively long exposure duration (greater than several hundred ms). The present results show that segmentation based on X-junctions requires an exposure duration of only about 40 ms (including mask duration). This short duration is consistent with the case of ‘static’ X-junctions (60 ms, Watanabe and Cavanagh 1992). Why are X-junctions a strong cue for segmentation? I speculate that under natural dynamic conditions, binocular disparity and motion signals do not provide reliable information for segmentation, since the visual system cannot code the high-speed dynamic signal (Burr and Ross 1982), which is used for binocular matching and motion coding. On the other hand, same-polarity X-junctions are an intrinsic feature of temporal integration accompanied by high-speed motion; thus a mechanism sensitive to the junctions has an ecological advantage, even under dynamic conditions.

References

Adelson E H, Anandan P, 1990 "Ordinal characteristics of transparency" *AAAI-90 Workshop on Qualitative Vision* (Boston, MA) pp 77-81.

Akerstrom R A, Todd J T, 1988 "The perception of stereoscopic transparency" *Perception & Psychophysics* **44** 421-432

Anderson B L, 1997 "A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions" *Perception* **26** 419-453

Beaudot W H A, 2002 "Role of onset asynchrony in contour integration" *Vision Research* **42** 1-9

Beck J, Prazdny K, Ivry R, 1984 "The perception of transparency with achromatic colors" *Perception & Psychophysics* **35** 407-422

Bowen R W, Wilson H R, 1994 "A two-process analysis of pattern masking" *Vision Research* **34** 645-657

Breitmeyer B G, Ogmen H, 2000 "Recent models and findings in visual backward masking: a comparison, review, and update" *Perception & Psychophysics* **62** 1572-1595

Burr D C, Ross J, 1982 "Contrast sensitivity at high velocities" *Vision Research* **22** 479-484

Dixon P, Di Lollo V, 1994 "Beyond visible persistence: an alternative account of temporal integration and segregation in visual processing" *Cognitive Psychology* **26** 33-63

Enns J T, Di Lollo V, 1997 "Object substitution: a new form of masking in unattended visual locations" *Psychological Science* **8** 135-139

Forte J, Hogben J H, Ross J, 1999 "Spatial limitations of temporal segmentation" *Vision Research* **39** 4052-4061

Foley J M, Chen C C, 1999 "Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: threshold measurements and a model" *Vision Research* **39** 3855-3872

Gegenfurtner K R, Rieger J, 2000 "Sensory and cognitive contributions of color to the recognition of natural scenes" *Current Biology* **10** 805-808

Geisler W S, 1999 "Motion streaks provide a spatial code for motion direction" *Nature* **400** 65-69

Georgeson M A, 1988 "Spatial phase dependence and the role of motion detection in monocular and dichoptic forward masking" *Vision Research* **28** 1193-1205

Hartmann E, Lachenmayr B, Brettel H, 1979 "The peripheral critical flicker frequency" *Vision Research* **19** 1019-1023

Holcombe A O, 2001 "A purely temporal transparency mechanism in the visual system" *Perception* **30** 1311-1320

Jiang Y, Chun M M, 2001 "The spatial gradient of visual masking by object substitution" *Vision Research* **41** 3121-3131

Julesz B, 1991 "Early vision and focal attention" *Reviews of Modern Physics* **63** 735-772

Kersten D, 1991 "Transparency and the cooperative computation of scene attributes", in *Computational Models of Visual Processing* Eds M S Landy, J A Movshon (Cambridge, MA: MIT Press) pp 209-228

Masson G S, Mestre D R, Stone L S, 1999 "Speed tuning of motion segmentation and discrimination" *Vision Research* **39** 4297-4308

Metelli F, 1974 “The perception of transparency” *Scientific American* **230** 90-98

Mitsudo H, 2003 “Information regarding structure and lightness based on phenomenal transparency influences the efficiency of visual search” *Perception* **32** 53-66

Motoyoshi I, Nishida S, 2001 “Temporal resolution of orientation-based texture segregation” *Vision Research* **41** 2089-2105

Neri P, Heeger D J, 2002 “Spatiotemporal mechanisms for detecting and identifying image features in human vision” *Nature Neuroscience* **5** 812-816

Polat U, Sagi D, 1993 “Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments” *Vision Research* **33** 993-999

Rauschenberger R, Yantis S, 2001 “Masking unveils pre-modal completion representation in visual search” *Nature* **410** 369-372

Sagi D, Julesz B, 1985 “‘Where’ and ‘what’ in vision” *Science* **228** 1217-1219

Scialfa C T, Joffe K M, 1995 “Preferential processing of target features in texture segmentation” *Perception & Psychophysics* **57** 1201-1208

Shelley-Tremblay J, Mack A, 1999 “Metacontrast masking and attention” *Psychological Science* **10** 508-515

Singh M, Anderson B L, 2002 “Toward a perceptual theory of transparency” *Psychological Review* **109** 492-519

Solomon J A, Morgan M J, 2000 “Facilitation from collinear flanks is cancelled by non-collinear flanks” *Vision Research* **40** 279-286

Tanaka Y, Sagi D, 1998 “Long-lasting, long-range detection facilitation” *Vision Research* **38** 2591-2599

Watanabe T, Cavanagh P, 1992 “The role of transparency in perceptual grouping and pattern recognition” *Perception* **21** 133-139

Watanabe T, Cavanagh P, 1993 “Transparent surfaces defined by implicit X junctions” *Vision Research* **33** 2339-2346

Watson A B, Nachmias J, 1977 “Patterns of temporal integration in the detection of gratings” *Vision Research* **17** 893-902

Williams C B, Hess R F, 1998 “Relationship between facilitation at threshold and suprathreshold contour integration” *Journal of the Optical Society of America A* **15** 2046-2051

Wolf W, Deubel H, 1997 “P31 phosphor persistence at photopic mean luminance level” *Spatial Vision* **10** 323-333

Wolfe J M, 1992 “‘Effortless’ texture segmentation and ‘parallel’ visual search are *not* the same thing” *Vision Research* **32** 757-763

Xing J, Heeger D J, 2000 “Center-surround interactions in foveal and peripheral vision” *Vision Research* **40** 3065-3072

Author's Note

Hiroyuki Mitsudo, Graduate School of Human-Environment Studies, Kyushu University.

A portion of this study was presented at the Second Asian Conference on Vision in Gyeongju, Korea, July 2002.

I thank Dr. Sachio Nakamizo, Takako Saitou and Takahiro Kawabe for their thoughtful comments on earlier versions of this manuscript and for helpful discussions regarding this study. I also thank Dr. Shoji Sunaga and Daiichiro Kuroki for their advice and assistance in measuring the phosphor persistence of the monitor with an oscilloscope. I am also grateful to Miss Katherine Miller (Royal English Language centre, Fukuoka, Japan) for revising the English used in this manuscript. I greatly thank the two anonymous reviewers for their helpful comments on earlier versions of this manuscript.

Correspondence should be addressed to Hiroyuki Mitsudo at the Department of Psychology, Faculty of Letters, Kyushu University, 6-19-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. Electronic mail may be sent to mitsudo@psycho.hes.kyushu-u.ac.jp.

Footnotes

[1] Hereafter, when referring to such ‘three-dimensional’ segmentation of stratified surfaces, I use the term ‘surface stratification’, distinguished from image (or two-dimensional) segmentation.

[2] Although Wolfe (1992) showed that data from visual-search and texture-segmentation experiments do not reflect exactly the same parallel processing, the visual search paradigm is also useful for revealing the process for segmenting a target that appears at an unpredictable location within the visual field.

[3] In static scenes, some same-polarity X-junctions can be formed by a shadow cast on a complex background. Such junctions, however, are not typical same-polarity X-junctions. This is because cast shadows always accompany a decrease in luminance; however, a transparent surface does not always accompany it. For this reason I do not offer any further discussion on cast shadows.

[4] Within the above-mentioned framework, a transparency judgment can be conceptually distinguished from image segmentation. For example, two-dimensional segmentation without transparency is fairly possible as a ‘mosaic’ surface without stratification (as shown in Figure 1B), by calculating luminance differences in the image.

[5] The shortest exposure duration (13 ms for each of the target and mask frames and an inter-stimulus interval of 13 ms) corresponds to an alternation rate of 25 Hz. Since this rate was below the critical flicker frequency at its mean luminance level (about 40 Hz, Hartmann et al 1979), the temporal transparency examined here was different from a by-product of simple flicker fusion.

[6] One might argue that static cues for transparency are not completely removed since the target frame contains T-junctions, some of which can trigger transparency (Watanabe and Cavanagh 1993). According to the previous study, transparency occurs when the luminance of an occluding region is not intermediate between those of two partly occluded regions. Since the stimuli used in the present study violated this constraint on transparency, static cues for transparency were completely removed in the stimuli.

[7] According to Wolf and Deubel (1997), under photopic conditions (mean luminance, about 20 cd/m^2), residual contrast due to phosphor persistence is reduced to less than 5% of the original pattern within 10 ms after its stimulus offset. In the present study, no occurrence of optical summation on the monitor was confirmed by recording its temporal luminance change using a phototransistor (TOSHIBA TPS603A) and an oscilloscope (HITACHI V-302B 30MHz).

[8] It could be thought that it is inappropriate to compare frequencies which had unequal chance levels. I thus corrected the chance levels for the two polarity conditions to the same level, by transforming these frequencies into probabilities under the condition in which the correct target location was reported. Even in this case, probabilities of the ‘transparent’ responses in the opposite-polarity condition were lower than (or almost identical to) those of the same-polarity condition. Accordingly, the transparency observed in the SOAs is not a response bias.

Figure Captions

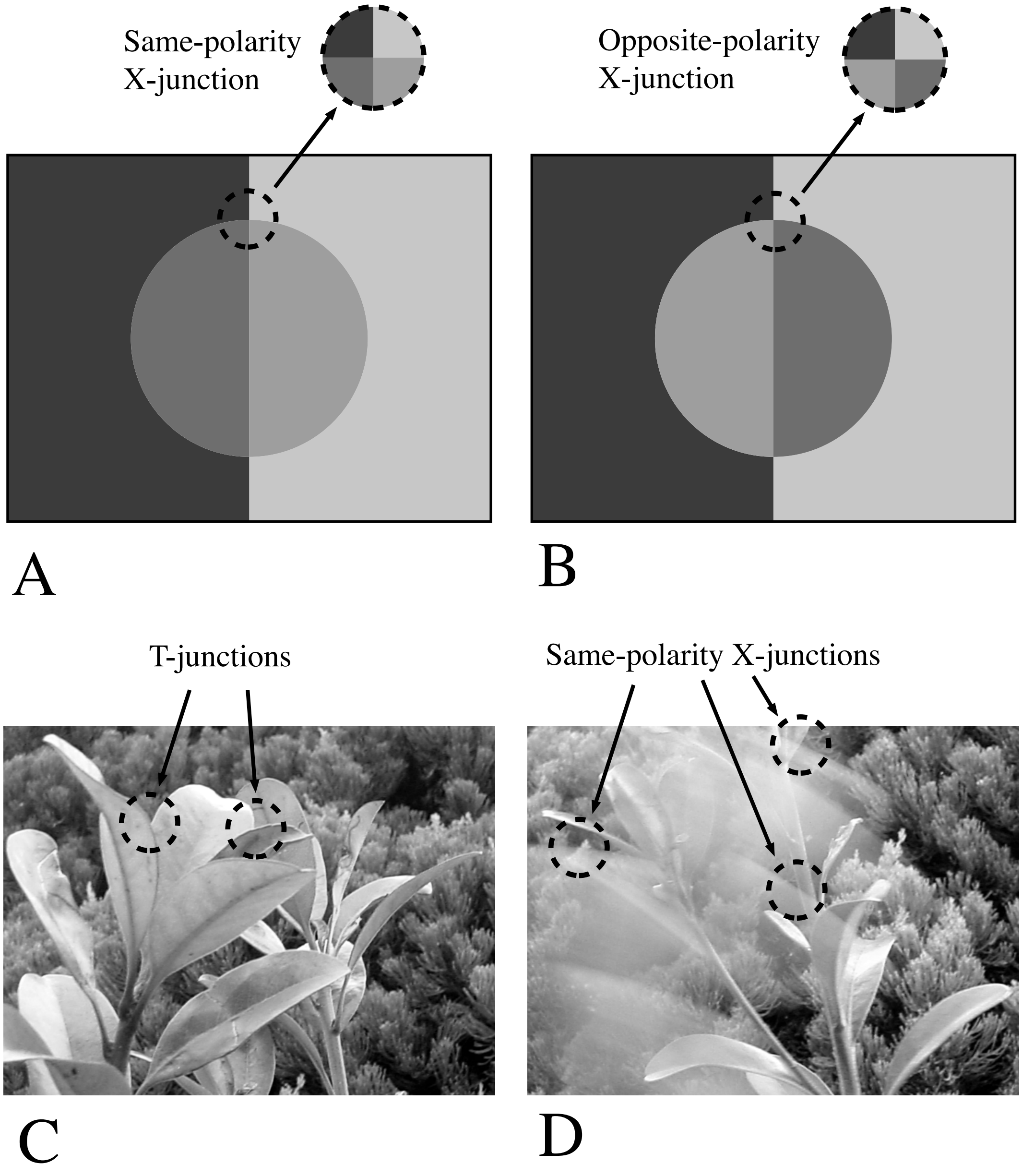
Figure 1. Perceptual transparency and junctions. A: A classical stimulus configuration for perceiving transparency contains same-polarity X-junctions. B: Opposite-polarity X-junctions do not trigger transparency. C: A natural static scene usually contains T-junctions for occlusion but not X-junctions. D: A dynamic version of the static scene (C) contains same-polarity X-junctions when luminance is temporally integrated (or averaged).

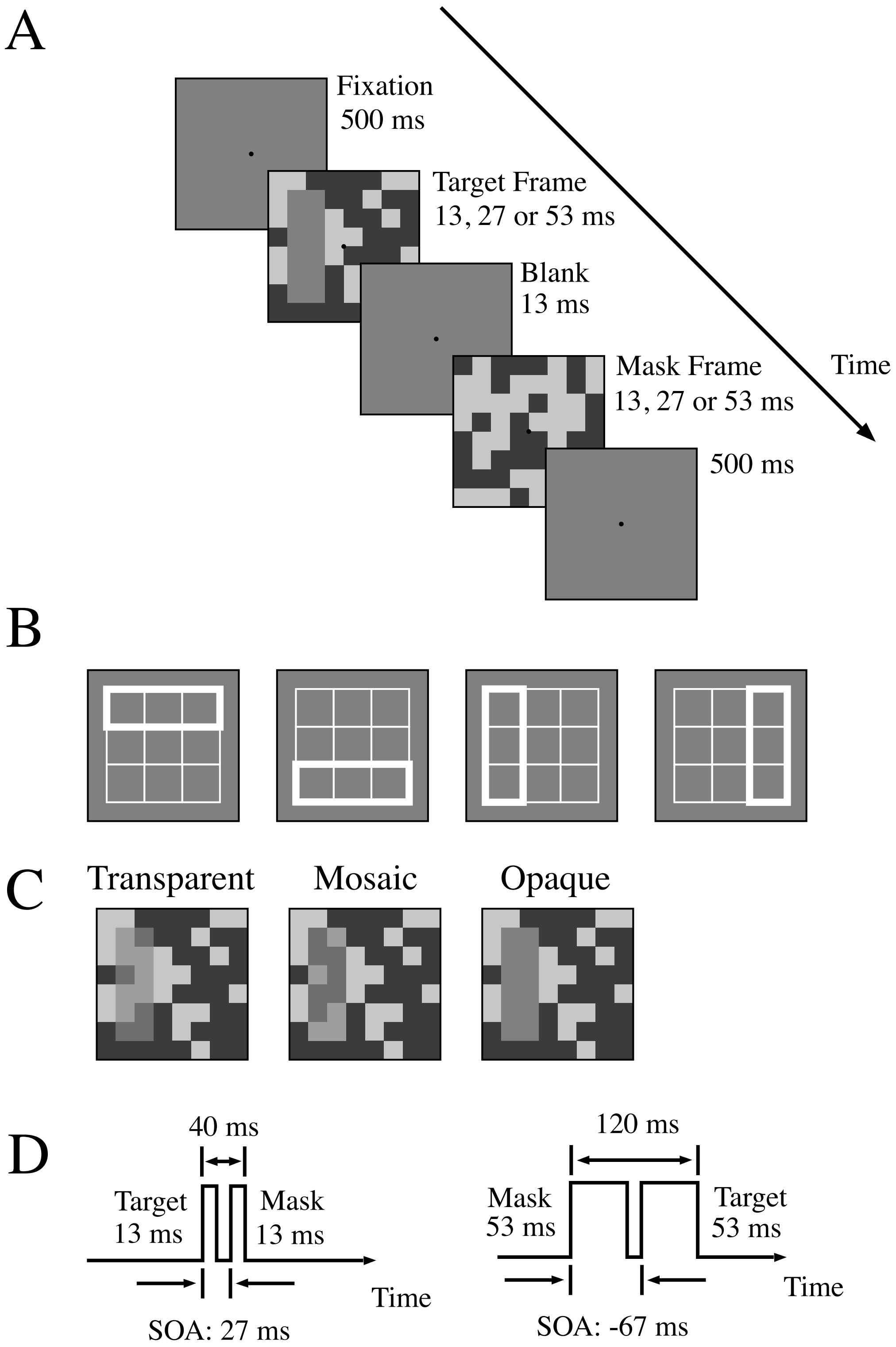
Figure 2. A: A schematic representation of the observation sequence. B: The four possible target locations for the segmentation task. C: An example of the response display for the identification task. D: Examples of the time course.

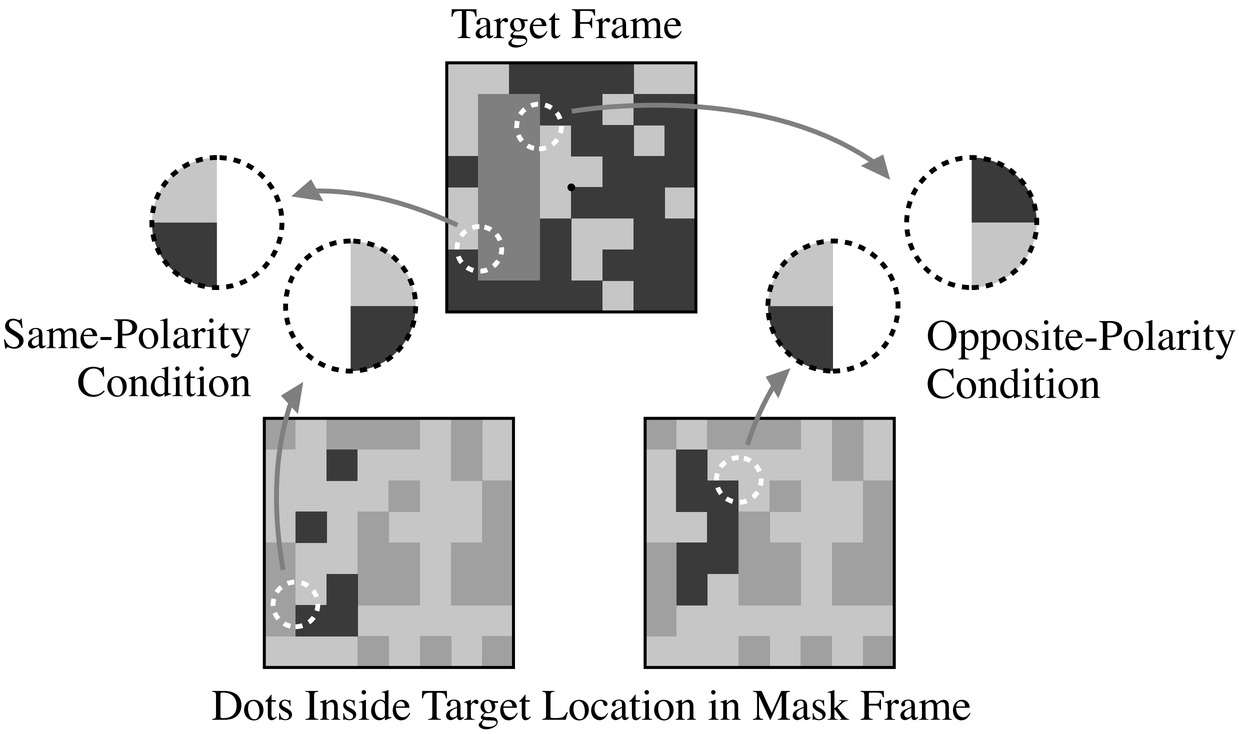
Figure 3. Contrast polarity at asynchronous X-junctions, in which static cues for both segmentation and transparency are eliminated. For the sake of clarity, dots appearing outside the target location in the mask frame are depicted in low contrast.

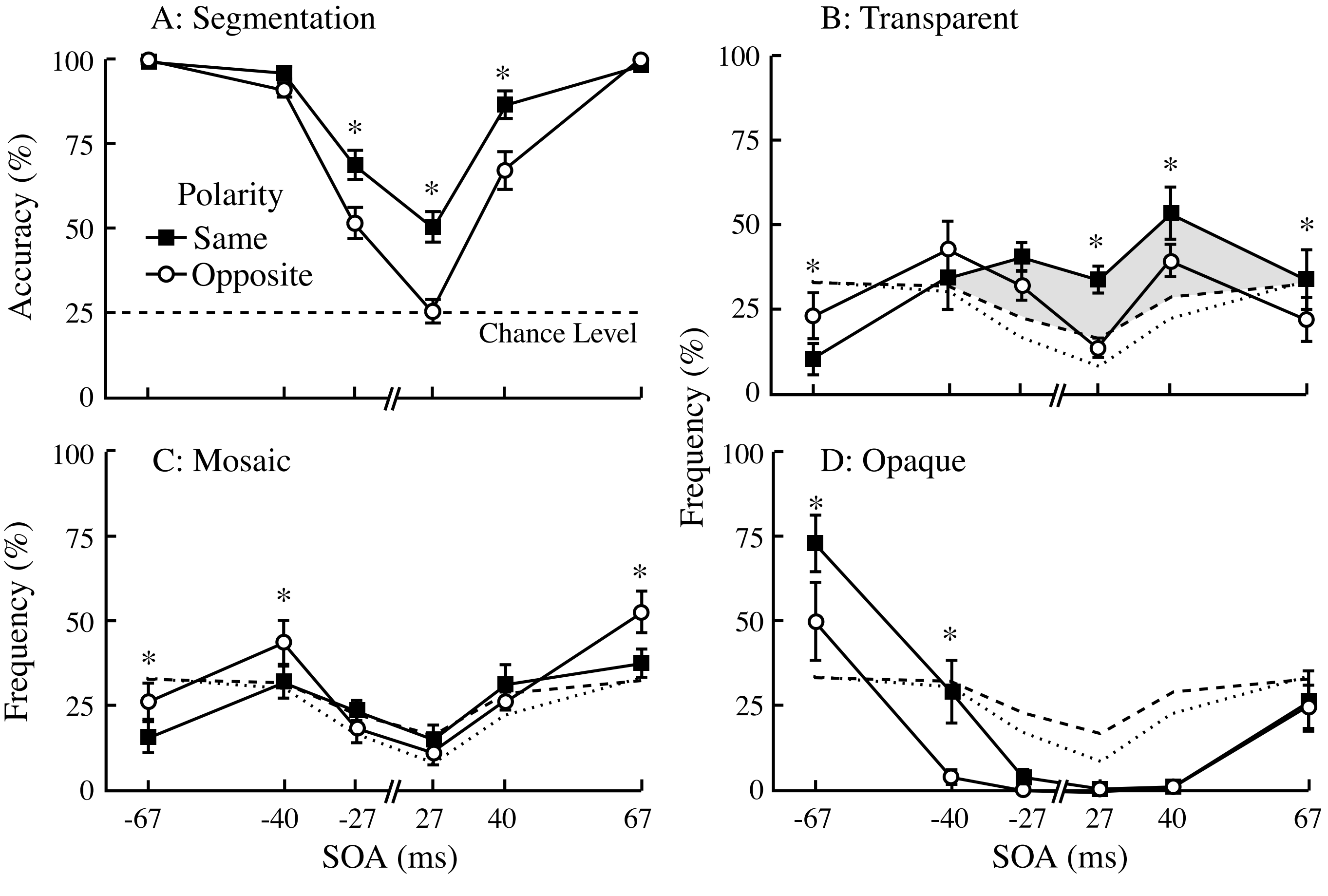
Figure 4. The results for the foveal group. A: The mean segmentation accuracy as a function of SOA. B, C and D: The mean correct frequency of each response in the identification task (transparent, mosaic and opaque, respectively) as a function of SOA. The sum of the three responses is equal to the mean segmentation accuracy for each polarity condition. Dashed and dotted lines indicate the chance levels for the same- and opposite-polarity conditions, respectively. The dark area in Figure 4B highlights the conditions in which transparency was observed. Vertical bars represent standard error. Asterisks indicate significant differences between the same- and opposite-polarity conditions (at least $p < .05$).

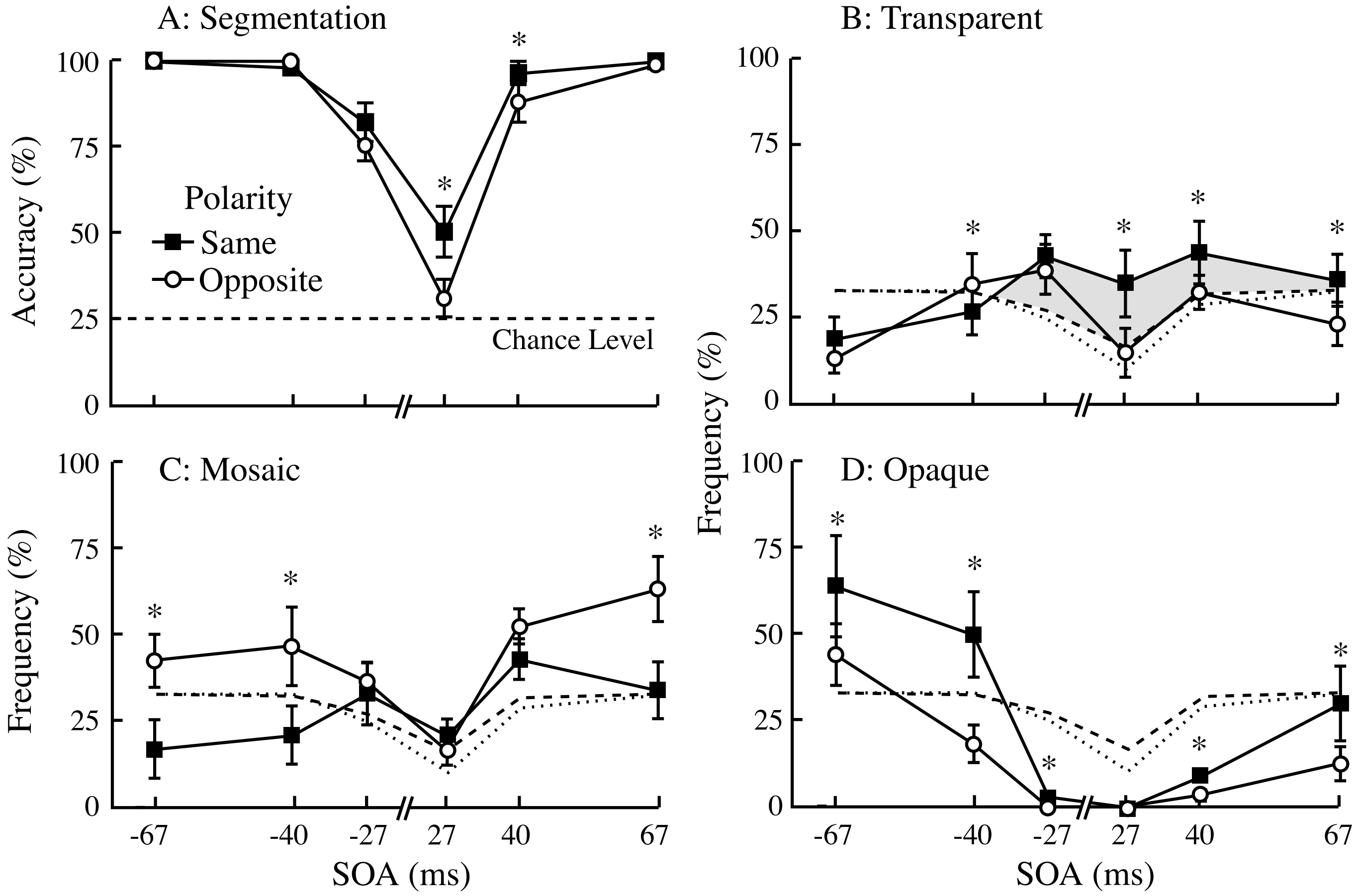
Figure 5. The results for the perifoveal group. A: The mean segmentation accuracy as a function of SOA. B, C and D: The mean correct frequency of each response in the identification task (transparent, mosaic and opaque, respectively) as a function of SOA. The sum of the three responses is equal to the mean segmentation accuracy for each polarity condition. Dashed and dotted lines indicate the chance levels for the same- and opposite-polarity conditions, respectively. The dark area in Figure 5B highlights the conditions in which transparency was observed. Vertical bars represent standard error. Asterisks indicate significant differences between the same- and opposite-polarity conditions (at least $p < .05$).


Same-polarity
X-junction
Opposite-polarity
X-junction
A
B
T-junctions
Same-polarity X-junctions
C
D


A
Fixation
500 ms
Target Frame
13, 27 or 53 ms
Blank
13 ms
Mask Frame
13, 27 or 53 ms
500 ms
Time
B
C
Transparent
Mosaic
Opaque
D
40 ms
Target
13 ms
Mask
13 ms
Time
SOA: 27 ms
120 ms
Mask
53 ms
Target
53 ms
Time
SOA: -67 ms


Target Frame
Same-Polarity
Condition
Opposite-Polarity
Condition
Dots Inside Target Location in Mask Frame


A: Segmentation
B: Transparent
C: Mosaic
D: Opaque
Accuracy (%)
Frequency (%)
Frequency (%)
100
75
50
25
0
Polarity
Same
Opposite
Chance Level
*
-67
-40
-27
27
40
67
SOA (ms)
SOA (ms)


A: Segmentation
B: Transparent
C: Mosaic
D: Opaque
Accuracy (%)
Frequency (%)
Frequency (%)
Polarity
Same
Opposite
Chance Level
100
75
50
25
0
*
-67
-40
-27
27
40
67
SOA (ms)
SOA (ms)